



# THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **Genome-wide by environment interaction studies of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland**

#### **Citation for published version:**

Arnau soler, A, MacDonald-Dunlop, E, Adams, MJ, Clarke, T-K, MacIntyre, D, Milburn, K, Navrady, L, Scotland, G, Consortium, MDDWGOTPG, Hayward, C, McIntosh, A & Thomson, P 2019, 'Genome-wide by environment interaction studies of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland' *Translational Psychiatry*, vol. 9, no. 1, 14. DOI: 10.1038/s41398-018-0360-y

#### **Digital Object Identifier (DOI):**

[10.1038/s41398-018-0360-y](https://doi.org/10.1038/s41398-018-0360-y)

#### **Link:**

[Link to publication record in Edinburgh Research Explorer](#)

#### **Document Version:**

Publisher's PDF, also known as Version of record

#### **Published In:**

*Translational Psychiatry*

#### **General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

#### **Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



ARTICLE

Open Access

# Genome-wide by environment interaction studies of depressive symptoms and psychosocial stress in UK Biobank and Generation Scotland

Aleix Arnau-Soler<sup>1</sup>, Erin Macdonald-Dunlop<sup>2</sup>, Mark J. Adams<sup>3</sup>, Toni-Kim Clarke<sup>3</sup>, Donald J. MacIntyre<sup>3</sup>, Keith Milburn<sup>4</sup>, Lauren Navrady<sup>3</sup>, Generation Scotland<sup>5</sup>, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Caroline Hayward<sup>6</sup>, Andrew M. McIntosh<sup>3</sup> and Pippa A. Thomson<sup>1</sup>

## Abstract

Stress is associated with poorer physical and mental health. To improve our understanding of this link, we performed genome-wide association studies (GWAS) of depressive symptoms and genome-wide by environment interaction studies (GWEIS) of depressive symptoms and stressful life events (SLE) in two UK population-based cohorts (Generation Scotland and UK Biobank). No SNP was individually significant in either GWAS, but gene-based tests identified six genes associated with depressive symptoms in UK Biobank (*DCC*, *ACSS3*, *DRD2*, *STAG1*, *FOXP2* and *KYNU*;  $p < 2.77 \times 10^{-6}$ ). Two SNPs with genome-wide significant GxE effects were identified by GWEIS in Generation Scotland: rs12789145 (53-kb downstream *PIWIL4*;  $p = 4.95 \times 10^{-9}$ ; total SLE) and rs17070072 (intronic to *ZCCHC2*;  $p = 1.46 \times 10^{-8}$ ; dependent SLE). A third locus upstream *CYLC2* (rs12000047 and rs12005200,  $p < 2.00 \times 10^{-8}$ ; dependent SLE) when the joint effect of the SNP main and GxE effects was considered. GWEIS gene-based tests identified: *MTNR1B* with GxE effect with dependent SLE in Generation Scotland; and *PHF2* with the joint effect in UK Biobank ( $p < 2.77 \times 10^{-6}$ ). Polygenic risk scores (PRSs) analyses incorporating GxE effects improved the prediction of depressive symptom scores, when using weights derived from either the UK Biobank GWAS of depressive symptoms ( $p = 0.01$ ) or the PGC GWAS of major depressive disorder ( $p = 5.91 \times 10^{-3}$ ). Using an independent sample, PRS derived using GWEIS GxE effects provided evidence of shared aetiologies between depressive symptoms and schizotypal personality, heart disease and COPD. Further such studies are required and may result in improved treatments for depression and other stress-related conditions.

## Introduction

Mental illness results from the interplay between genetic susceptibility and environmental risk factors<sup>1,2</sup>. Previous studies have shown that the effects of environmental factors on traits may be partially heritable<sup>3</sup> and moderated by genetics<sup>4,5</sup>. Major depressive disorder (MDD) is the most common psychiatric disorder with a lifetime prevalence of approximately 14% globally<sup>6</sup> and with a heritability of approximately 37%<sup>7</sup>. There is strong evidence for the role of stressful life events (SLEs) as risk factor and trigger for

Correspondence: Aleix Arnau-Soler (aleix.arnau.soler@igmm.ed.ac.uk) or Pippa A. Thomson (Pippa.Thomson@ed.ac.uk)

<sup>1</sup>Medical Genetics Section, University of Edinburgh, Centre for Genomic and Experimental Medicine and MRC Institute of Genetics and Molecular Medicine, Edinburgh, UK

<sup>2</sup>Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, UK Full list of author information is available at the end of the article.

Generation Scotland is a collaboration between the University Medical School and NHS in Aberdeen, Dundee, Edinburgh and Glasgow, Scotland, UK

© The Author(s) 2019



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

depression<sup>8–12</sup>. Genetic control of sensitivity to stress may vary between individuals, resulting in individual differences in the depressogenic effects of SLE, i.e., genotype-by-environment interaction (GxE)<sup>4,13–16</sup>. Significant evidence of GxE has been reported for common respiratory diseases and some forms of cancer<sup>17–22</sup>, and GxE studies have identified genetic risk variants not found by genome-wide association studies (GWAS)<sup>23–27</sup>.

Interaction between polygenic risk of MDD and recent SLE are reported to increase liability to depressive symptoms<sup>4,16</sup>; validating the implementation of genome-wide approaches to study GxE in depression. Most GxE studies for MDD have been conducted on candidate genes, or using polygenic approaches to a wide range of environmental risk factors, with some contradictory findings<sup>28–32</sup>. Incorporating knowledge about recent SLE into GWAS may improve our ability to detect risk variants in depression otherwise missed in GWAS<sup>33</sup>. To date, three studies have performed genome-wide by environment interaction studies (GWEIS) of MDD and SLE<sup>34–36</sup>, but this is the first study to perform GWEIS of depressive symptoms using adult SLE in cohorts of relatively homogeneous European ancestry.

Interpretation of GxE effects may be hindered by gene–environment correlation. Gene–environment correlation denotes a genetic mediation of associations through genetic influences on exposure to, or reporting of, environments<sup>2,37</sup>. Genetic factors predisposing to MDD may contribute to exposure and/or reporting of SLE<sup>38</sup>. To tackle this limitation, measures of SLE can be broken down into SLE likely to be independent of a respondent's own behaviour and symptoms, or into dependent SLE, in which participants may play an active role exposure to SLE<sup>39,40</sup>. Different genetic influences, including a higher heritability, are reported for dependent SLE compared to independent SLE<sup>38,41–44</sup>, suggesting that whereas GxE driven by independent SLE is likely to reflect a genetic moderation of associations between SLE and depression, GxE driven by dependent SLE may result from a genetic mediation of the association through genetically driven personality or behavioural traits. To test this, we analysed dependent and independent SLE scores separately in Generation Scotland (GS).

Stress contributes to many human conditions, with evidence of genetic vulnerability to the effect of SLE<sup>45</sup>. Therefore, genetic stress-response factors in MDD may also underlie the aetiology of other stress-linked disorders with which MDD is often comorbid<sup>46,47</sup> (e.g., cardiovascular diseases<sup>48</sup>, diabetes<sup>49</sup>, chronic pain<sup>50</sup> and inflammation<sup>51</sup>). We tested the hypothesis that pleiotropy and shared aetiology between mental and physical health conditions may be due in part to genetic variants underlying SLE effects in depression.

In this study, we conduct GWEIS of depressive symptoms incorporating data on SLE in two independent UK-based cohorts. We aimed to: (i) identify loci associated with depressive symptoms through genetic response to SLE; (ii) study dependent and independent SLE to support a contribution of genetically mediated exposure to stress; (iii) assess whether GxE effects improve the proportion of phenotypic variance in depressive symptoms explained by genetic additive main effects alone; and (iv) test for a significant overlap in the genetic aetiology of the response to SLE and mental and physical stress-related phenotypes.

## Materials and methods

The core workflow of this study is summarised in Fig. 1.

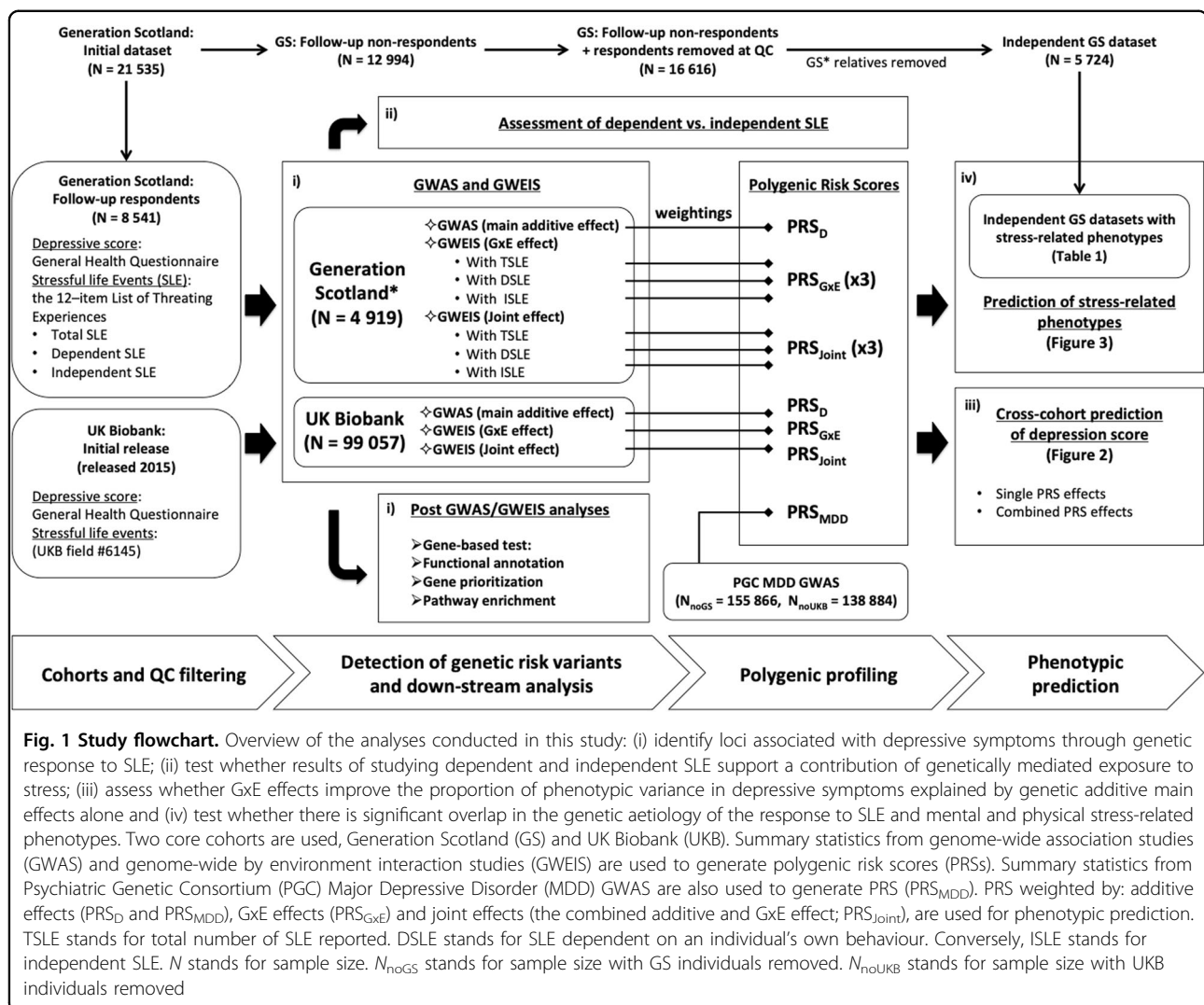
### Cohort descriptions

#### GS

GS is a family-based population cohort representative of the Scottish population<sup>52</sup>. At baseline, blood and salivary DNA samples were collected, stored and genotyped at the Wellcome Trust Clinical Research Facility, Edinburgh. Genome-wide genotype data were generated using the Illumina HumanOmniExpressExome-8 v1.0 DNA Analysis BeadChip (San Diego, CA, USA) and Infinium chemistry<sup>53</sup>. The procedures and details for DNA extraction and genotyping have been extensively described elsewhere<sup>54,55</sup>. In total, 21,525 participants were re-contacted to participate in a follow-up mental health study (Stratifying Resilience and Depression Longitudinally, STRADL), of which 8541 participants responded providing updated measures in psychiatric symptoms and SLE through self-reported mental health questionnaires<sup>56</sup>. Samples were excluded if: they were duplicate samples, had diagnoses of bipolar disorder, no SLE data (non-respondents), were population outliers (mainly non-Caucasians and Italian ancestry subgroup), had sex mismatches or were missing >2% of genotypes. Single nucleotide polymorphisms (SNPs) were excluded if: missing >2% of genotypes, Hardy–Weinberg equilibrium test  $p < 1 \times 10^{-6}$ , or minor allele frequency <1%. Further details of the GS and STRADL cohort are available elsewhere<sup>52,56–58</sup>. All components of GS and STRADL obtained ethical approval from the Tayside Committee on Medical Research Ethics on behalf of the NHS (reference 05/s1401/89). After quality control, individuals were filtered by degree of relatedness ( $\pi$ -hat < 0.05), maximising retention of those individuals reporting a higher number of SLE. The final dataset comprised data on 4919 unrelated individuals (1929 men; 2990 women) and 560,351 SNPs.

#### Independent GS datasets

Additional datasets for a range of stress-linked medical conditions and personality traits were created from GS ( $N$



**Fig. 1 Study flowchart.** Overview of the analyses conducted in this study: (i) identify loci associated with depressive symptoms through genetic response to SLE; (ii) test whether results of studying dependent and independent SLE support a contribution of genetically mediated exposure to stress; (iii) assess whether GxE effects improve the proportion of phenotypic variance in depressive symptoms explained by genetic additive main effects alone and (iv) test whether there is significant overlap in the genetic aetiology of the response to SLE and mental and physical stress-related phenotypes. Two core cohorts are used, Generation Scotland (GS) and UK Biobank (UKB). Summary statistics from genome-wide association studies (GWAS) and genome-wide by environment interaction studies (GWEIS) are used to generate polygenic risk scores (PRSs). Summary statistics from Psychiatric Genetic Consortium (PGC) Major Depressive Disorder (MDD) GWAS are also used to generate PRS (PRS<sub>MDD</sub>). PRS weighted by: additive effects (PRS<sub>D</sub> and PRS<sub>MDD</sub>), GxE effects (PRS<sub>GxE</sub>) and joint effects (the combined additive and GxE effect; PRS<sub>Joint</sub>), are used for phenotypic prediction. TSLE stands for total number of SLE reported. DSLE stands for SLE dependent on an individual's own behaviour. Conversely, ISLE stands for independent SLE. *N* stands for sample size. *N*<sub>noGS</sub> stands for sample size with GS individuals removed. *N*<sub>noUKB</sub> stands for sample size with UKB individuals removed

= 21,525) excluding respondents and their relatives (*N* = 5724). Following the same quality control criteria detailed above, we maximised unrelated non-respondents for retention of cases, or proxy cases (see below), to maximise the information available for each phenotype. This resulted in independent datasets with unrelated individuals for each trait. Differences between respondents and non-respondents are noted in the figure legend of Table 1.

#### UK Biobank (UKB)

This study used data from 99,057 unrelated individuals (47,558 men; 51,499 women) from the initial release of UKB genotyped data (released 2015; under UKB project 4844). Briefly, participants were removed based on UKB genomic analysis exclusion, non-white British ancestry, high missingness, genetic relatedness (kinship coefficient > 0.0442), QC failure in UK BiLEVE study and gender mismatch. GS participants and their relatives were

excluded and GS SNPs imputed to a reference set combining the UK10K haplotype and 1000 Genomes Phase 3 reference panels<sup>59</sup>. After quality control, 1,009,208 SNPs remained. UKB received ethical approval from the NHS National Research Ethics Service North West (reference: 11/NW/0382). Further details on UKB cohort description, genotyping, imputation and quality control are available elsewhere<sup>60–62</sup>.

All participants provided informed consent.

#### Phenotype assessment

##### SLEs

GS participants reported SLE experienced over the preceding 6 months through a self-reported brief life events questionnaire based on the 12-item list of threatening experiences<sup>39,63,64</sup> (Supplementary Table 1a). The total number of SLE reported (TSLE) consisted of the number of 'yes' responses. TSLE were subdivided into SLE

**Table 1 GS samples with stress-related phenotypes**

Trait	N	Males/females	N SNPs	N Cases	N Controls
Alzheimer (R)	3377	1475/1903	560,622	655	2722
Asthma	3390	1500/1890	560,569	555	2835
Asthma (R)	3375	1470/1905	560,432	910	2465
Bowel cancer (R)	3386	1495/1891	560,630	672	2714
Breast cancer	3388	1486/1902	560,611	83	3305
Breast cancer (R)	3386	1482/1904	560,579	564	2822
Chronic obstructive pulmonary disease	3387	1496/1891	560,591	73	3314
Chronic obstructive pulmonary disease (R)	3387	1474/1913	560,620	553	2834
Depression	3385	1495/1890	560,584	483	2902
Depression (R)	3382	1506/1876	560,514	731	2651
Diabetes	3388	1497/1891	560,469	185	3203
Diabetes (R)	3389	1481/1908	560,584	1144	2245
Heart disease	3392	1504/1888	560,526	212	3180
Heart disease (R)	3377	1483/1894	560,479	2254	1123
High blood pressure	3402	1501/1901	560,508	729	2673
High blood pressure (R)	3372	1464/1908	560,569	1901	1471
Hip fracture (R)	3388	1489/1899	560,572	421	2967
Lung cancer (R)	3379	1492/1887	560,600	798	2581
Osteoarthritis	3395	1486/1909	560,640	411	2984
Osteoarthritis (R)	3383	1466/1917	560,516	961	2422
Parkinson (R)	3388	1488/1900	560,590	236	3152
Prostate cancer (R)	3381	1495/1886	560,570	329	3052
Rheumatoid arthritis	3387	1490/1897	560,618	93	3294
Rheumatoid arthritis (R)	3380	1487/1893	560,543	765	2615
Stroke	3387	1492/1895	560,613	81	3306
Stroke (R)	3385	1463/1922	560,478	1506	1879
Neuroticism <sup>a</sup>	3421	1521/1900	560,484	-	-
Extraversion <sup>a</sup>	3420	1520/1900	560,476	-	-
Schizotypal personality <sup>a</sup>	2386	1065/1321	560,369	-	-
Mood disorder <sup>a</sup>	2307	1040/1267	560,318	-	-

Samples were maximised for retention of cases to maximise the information available for each trait. There was no preferential selection of relatives in pairs for quantitative phenotypes, in order to retain the underlying distribution. All individuals involved in the datasets listed above were non-respondents to the GS follow-up study. Compared with individuals included at GS GWEIS (respondents in GS follow-up), non-respondents were significantly: younger, from more socioeconomically deprived areas, generally less healthier and wealthier. Non-respondents were more likely to smoke, and less likely to drink alcohol, although they consumed more units per week, compared with respondents. At GS baseline, non-respondents experienced more psychological distress and reported higher scores in symptoms of GHQ-depression and GHQ-anxiety than respondents<sup>56</sup>

The total target sample size (N), number of males and females in N, number of SNPs (N SNPs) in target sample size N: the number of SNPs used as predictors after clumping step range between 90,650 and 91,000. The number of cases and controls in the independent target sample is indicated for binary phenotypes only. Samples were mapping by proxy approach was used (i.e., where first-degree relatives of individuals with the disease were considered proxy cases and included into the group of cases) are indicated by (R)

GS Generation Scotland, GWEIS genome-wide by environment interaction studies, GHQ General Health Questionnaire

<sup>a</sup>Assessed through self-reported questionnaires

potentially dependent or secondary to an individual's own behaviour (DSLE, questions 6–11 in Supplementary Table 1a), and independent SLE (ISLE, questions 1–5 in

Supplementary Table 1a; pregnancy item removed) following Brugha et al.<sup>39,40</sup>. Thus, three SLE measures (TSLE, DSLE and ISLE) were constructed for GS. UKB



participants were screened for ‘illness, injury, bereavement and stress’ (Supplementary Table 1b) over the previous 2 years using six items included in the UKB Touchscreen questionnaire. A score reflecting SLE reported in UKB (TSLE<sub>UKB</sub>) was constructed by summing the number of ‘yes’ responses.

### Psychological assessment

GS participants reported whether their current mental state over the preceding 2 weeks differed from their typical state using a self-administered 28-item scaled version of the General Health Questionnaire (GHQ)<sup>65–67</sup>. Participants rated the degree and severity of their current symptoms with a four-point Likert scale (following Goldberg et al.<sup>67</sup>). A final log-transformed GHQ was used to detect altered psychopathology and thus, assess depressive symptoms as results of SLE. In UKB participants, current depressive symptoms over the preceding 2 weeks were evaluated using four psychometric screening items (Supplementary Table 2), including two validated and reliable questions for screening depression<sup>68</sup>, from the Patient Health Questionnaire (PHQ) validated to screen mental illness<sup>69,70</sup>. Each question was rated in a four-point Likert scale to assess impairment/severity of symptoms. Due to its skewed distribution, a four-point PHQ score was formed from PHQ (0 = 0; 1 = 1–2; 2 = 3–5; 3 = 6 or more) to create a more normal distribution.

### Stress-related traits

Targeted GS stress-related phenotypes and sample sizes are shown in Table 1 and detailed elsewhere<sup>52</sup>. These conditions were selected from literature review based on previous evidence of a link with stress<sup>45</sup> (see also Supplementary Material: third section). Furthermore, we created additional independent samples using mapping by proxy, where individuals with a self-reported first-degree relative with a selected phenotype were included as proxy cases. This approach provides greater power to detect susceptibility variants in traits with low prevalence<sup>71</sup>.

### Statistical analyses

#### SNP-heritability and genetic correlation

A restricted maximum likelihood approach was applied to estimate SNP-heritability ( $h^2_{\text{SNP}}$ ) of depressive symptoms and self-reported SLE measures, and within samples bivariate genetic correlation between depressive symptoms and SLE measures using GCTA<sup>72</sup>.

#### GWAS analyses

GWAS were conducted in PLINK<sup>73</sup>. In GS, age, sex and 20 principal components (PCs) were fitted as covariates. In UKB, age, sex and 15 PCs recommended by UKB were fitted as covariates. The genome-wide significance threshold was  $p = 5 \times 10^{-8}$ .

#### GWEIS analyses

GWEIS were conducted on GHQ (the dependent variable) for TSLE, DSLE and ISLE in GS and on PHQ for TSLE<sub>UKB</sub> in UKB fitting the same covariates detailed above to reduce error variance. GWEIS were conducted using an R plugin for PLINK<sup>73</sup> developed by Almli et al.<sup>74</sup> (<https://epstein-software.github.io/robust-joint-interaction>). This method implements a robust test that jointly considers SNP and SNP–environment interaction effects from a full model ( $Y \sim \beta_0 + \beta_{\text{SNP}} + \beta_{\text{SLE}} + \beta_{\text{SNP} \times \text{SLE}} + \beta_{\text{Covariates}}$ ) against a null model where both the SNP and SNP×SLE effects equal 0, to assess the joint effect (the combined additive main and G×E genetic effect at a SNP) using a nonlinear statistical approach that applies Huber–White estimates of variance to correct possible inflation due to heteroscedasticity (unequal variances across exposure levels). This robust test should reduce confounding due to differences in variance induced by covariate interaction effects if present<sup>75</sup>. Additional code was added (courtesy of Prof. Michael Epstein;<sup>74</sup> Supplementary Material) to generate beta-coefficients and the  $p$ -value of the G×E term alone. In UKB, correcting for 1,009,208 SNPs and one exposure, we established a Bonferroni-adjusted threshold for significance at  $p = 2.47 \times 10^{-8}$  for both joint and G×E effects. In GS, correcting for 560,351 SNPs and three measures of SLE we established a genome-wide significance threshold of  $p = 2.97 \times 10^{-8}$ .

#### Post-GWAS/GWEIS analyses

GWAS and GWEIS summary statistics were analysed using FUMA<sup>76</sup> including: gene-based tests, functional annotation, gene prioritisation and pathway enrichment (Supplementary Material).

#### Polygenic profiling and prediction

Polygenic risk scores (PRSs) weighting by G×E effects (PRS<sub>G×E</sub>) were generated using PRSice-2<sup>77</sup> (Supplementary Material) in GS using G×E effects from UKB-GWEIS. In UKB, PRS<sub>G×E</sub> were constructed using G×E effects from all three GS-GWEIS (TSLE, DSLE and ISLE as exposures) independently. PRS were also weighted in both samples using either UKB-GWAS or GS-GWAS statistics (PRS<sub>D</sub>), and summary statistics from Psychiatric Genetic Consortium (PGC) MDD-GWAS (released 2016; PRS<sub>MDD</sub>) that excluded GS and UKB individuals when required ( $N_{\text{noGS}} = 155,866$ ;  $N_{\text{noUKB}} = 138,884$ ). Furthermore, we calculated PRS weighted by the joint effects (the combined additive main and G×E genetic effects; PRS<sub>joint</sub>) from either the UKB-GWEIS or GS-GWEIS. PRS predictions of depressive symptoms were permuted 10,000 times. Multiple regression models fitting PRS<sub>G×E</sub> and PRS<sub>MDD</sub>, and both PRS<sub>G×E</sub> and PRS<sub>D</sub> were tested. All models were adjusted by same covariates used in GWAS/

GWEIS. Null models were estimated from the direct effects of covariates alone. The predictive improvement of combining  $PRS_{GxE}$  and  $PRS_{MDD}/PRS_D$  effects over  $PRS_{MDD}/PRS_D$  effects alone was tested for significance using the likelihood ratio test (LRT).

Prediction of  $PRS_D$ ,  $PRS_{GxE}$  and  $PRS_{joint}$  of stress-linked traits were adjusted by age, sex and 20 PCs; and permuted 10,000 times. Empirical- $p$ -values after permutations were further adjusted by false discovery rate (FDR, conservative threshold at Empirical- $p = 6.16 \times 10^{-3}$ ). The predictive improvement of fitting  $PRS_{GxE}$  combined with  $PRS_D$  and covariates over prediction of a phenotype using the  $PRS_D$  effect alone with covariates was assessed using LRT, and LRT- $p$ -values adjusted by FDR (conservative threshold at LRT- $p = 8.35 \times 10^{-4}$ ).

## Results

### Phenotypic and genetic correlations

Depressive symptom scores and SLE measures were positively correlated in both UKB ( $r^2 = 0.22$ ,  $p < 2.2 \times 10^{-16}$ ) and GS (TSLE- $r^2 = 0.21$ ,  $p = 1.69 \times 10^{-52}$ ; DSLE- $r^2 = 0.21$ ,  $p = 8.59 \times 10^{-51}$ ; ISLE- $r^2 = 0.17$ ,  $p = 2.33 \times 10^{-33}$ ). Significant bivariate genetic correlation between depression and SLE scores was identified in UKB ( $r_G = 0.72$ ;  $p < 1 \times 10^{-5}$ ,  $N = 50,000$ ), but not in GS ( $r_G = 1$ ,  $p = 0.056$ ,  $N = 4919$ ; Supplementary Table 3a).

### SNP-heritability ( $h^2_{SNP}$ )

In UKB, a significant  $h^2_{SNP}$  of PHQ was identified ( $h^2_{SNP} = 0.090$ ;  $p < 0.001$ ;  $N = 99,057$ ). This estimate remained significant after adjusting by TSLE<sub>UKB</sub> effect ( $h^2_{SNP} = 0.079$ ;  $p < 0.001$ ), suggesting a genetic contribution unique to depressive symptoms. The  $h^2_{SNP}$  of TSLE<sub>UKB</sub> was also significant ( $h^2_{SNP} = 0.040$ ,  $p < 0.001$ ; Supplementary Table 3b). In GS,  $h^2_{SNP}$  was not significant for GHQ ( $h^2_{SNP} = 0.071$ ,  $p = 0.165$ ;  $N = 4919$ ). However, in an *ad hoc* estimation from the baseline sample of 6751 unrelated GS participants (details in Supplementary Table 3b) we detected a significant  $h^2_{SNP}$  for GHQ ( $h^2_{SNP} = 0.135$ ;  $p < 5.15 \times 10^{-3}$ ), suggesting that the power to estimate  $h^2_{SNP}$  in GS may be limited by sample size. Estimates were not significant for either TSLE ( $h^2_{SNP} = 0.061$ ,  $p = 0.189$ ; Supplementary Table 3b) or ISLE ( $h^2_{SNP} = 0.000$ ,  $p = 0.5$ ), but  $h^2_{SNP}$  was significant for DSLE ( $h^2_{SNP} = 0.131$ ,  $p = 0.029$ ), supporting a potential genetic mediation and gene–environment correlation.

### GWAS of depressive symptoms

No genome-wide significant SNPs were detected by GWAS in either cohort. Top findings ( $p < 1 \times 10^{-5}$ ) are summarised in Supplementary Table 4. Manhattan and QQ plots are shown in Supplementary Figures 1–4. There was no evidence of genomic inflation (all  $\lambda_{1000} < 1.01$ ).

### Post-GWAS analyses

Gene-based test identified six genes associated with PHQ using the UKB-GWAS statistics at genome-wide significance (Bonferroni-corrected  $p = 2.77 \times 10^{-6}$ ; *DCC*,  $p = 7.53 \times 10^{-8}$ ; *ACSS3*,  $p = 6.51 \times 10^{-7}$ ; *DRD2*,  $p = 6.55 \times 10^{-7}$ ; *STAG1*,  $p = 1.63 \times 10^{-6}$ ; *FOXP2*,  $p = 2.09 \times 10^{-6}$ ; *KYNU*,  $p = 2.24 \times 10^{-6}$ ; Supplementary Figure 8). Prioritised genes based on position, expression quantitative trait loci (eQTL) and chromatin interaction mapping are detailed in Supplementary Table 5. No genes were detected in GS-GWAS gene-based test (Supplementary Figures 9). No tissue-specific enrichment was detected from GWAS in either cohort. Significant gene-sets and GWAS catalogue associations for UKB-GWAS are reported in Supplementary Table 6. These included the *biological process*: positive regulation of long-term synaptic potentiation, and *GWAS catalogue associations*: brain structure, schizophrenia, response to amphetamines, age-related cataracts (age at onset), fibrinogen, acne (severe), fibrinogen levels and educational attainment; all adjusted- $p < 0.01$ . There was no significant gene-set enrichment from GS-GWAS.

### GWEIS of depressive symptoms

Manhattan and QQ plots are shown in Supplementary Figures 1–4. There was no evidence of GWEIS inflation for either UKB or GS (all  $\lambda_{1000} < 1.01$ ). No genome-wide significant GWEIS associations were detected for SLE in UKB. GS-GWEIS using TSLE identified a significant GxE effect ( $p < 2.97 \times 10^{-8}$ ) at an intragenic SNP on chromosome 11 (rs12789145,  $p = 4.95 \times 10^{-9}$ ,  $\beta = 0.06$ , closest gene: *PIWIL4*; Supplementary Figure 5), and using DSLE at an intronic SNP in *ZCCHC2* on chromosome 18 (rs17070072,  $p = 1.46 \times 10^{-8}$ ,  $\beta = -0.08$ ; Supplementary Figure 6). In their corresponding joint effect tests, both rs12789145 ( $p = 2.77 \times 10^{-8}$ ) and rs17070072 ( $p = 1.96 \times 10^{-8}$ ) were significant. GWEIS for joint effect using DSLE identified two further significant SNPs on chromosome 9 (rs12000047,  $p = 2.00 \times 10^{-8}$ ,  $\beta = -0.23$ ; rs12005200,  $p = 2.09 \times 10^{-8}$ ,  $\beta = -0.23$ , LD  $r^2 > 0.8$ , closest gene: *CYLC2*; Supplementary Figure 7). None of these associations replicated in UKB ( $p > 0.05$ ), although the effect direction was consistent between cohorts for the SNP close to *PIWL1* and SNPs at *CYLC2*. No SNP achieved genome-wide significant association in the GS-GWEIS using ISLE as exposure. Top GWEIS results ( $p < 1 \times 10^{-5}$ ) are summarised in Supplementary Tables 7–10.

### Post-GWEIS analyses: gene-based tests

All results are shown in Supplementary Figures 10–17. Two genes were associated with PHQ using the joint effect from the UKB-GWEIS (*ACSS3*  $p = 1.61 \times 10^{-6}$ ; *PHF2*,  $p = 2.28 \times 10^{-6}$ ; Supplementary Figure 11). *ACSS3* was previously identified using the additive main effects,

whereas *PHF2* was only significantly associated using the joint effects. Gene-based tests identified *MTNR1B* as significantly associated with GHQ on the GS-GWEIS using DSLE in both GxE ( $p = 1.53 \times 10^{-6}$ ) and joint effects ( $p = 2.38 \times 10^{-6}$ ; Supplementary Figures 14–15).

### Post-GWEIS analyses: tissue enrichment

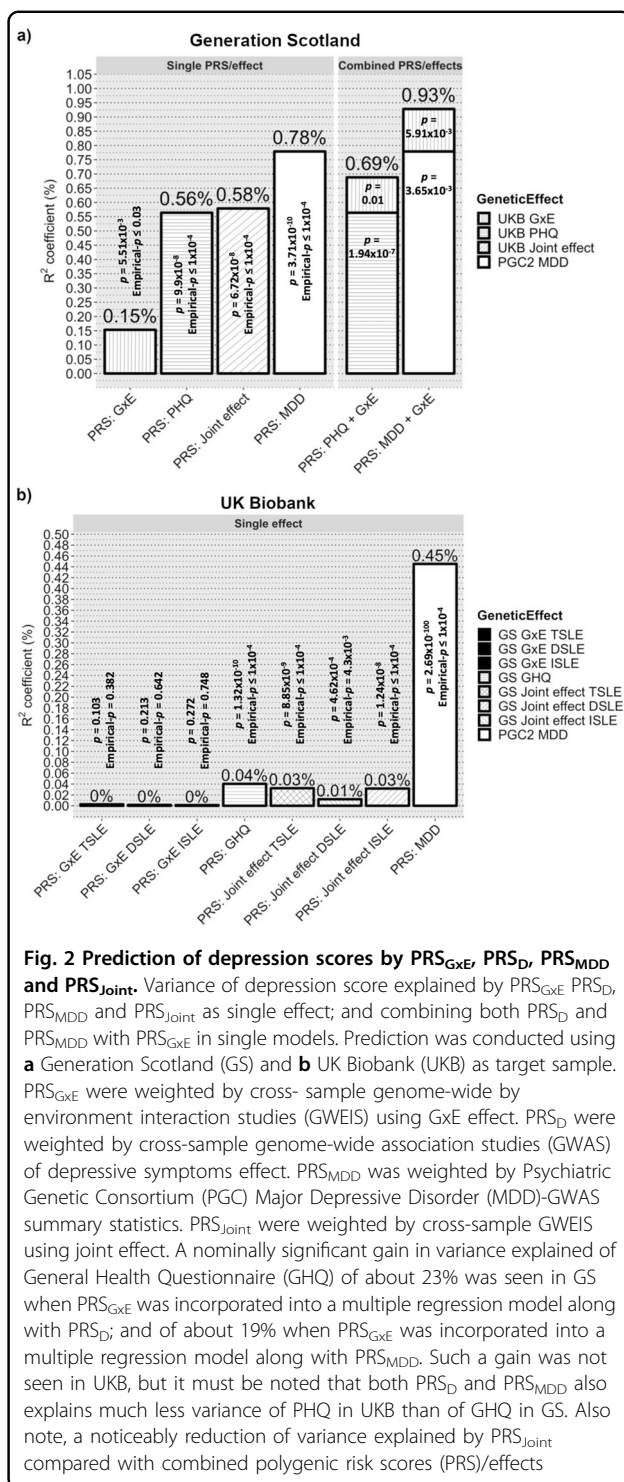
We prioritised genes based on position, eQTL and chromatin interaction mapping in brain tissues and regions. In UKB, prioritised genes using GxE effects were enriched for upregulated differentially expressed genes from adrenal gland (adjusted- $p = 3.58 \times 10^{-2}$ ). Using joint effects, prioritised genes were enriched on upregulated differentially expressed genes from artery tibial (adjusted- $p = 4.34 \times 10^{-2}$ ). In GS, prioritised genes were enriched: in upregulated differentially expressed genes from artery coronary (adjusted- $p = 4.55 \times 10^{-2}$ ) using GxE effects with DSLE; in downregulated differentially expressed genes from artery aorta tissue (adjusted- $p = 4.71 \times 10^{-2}$ ) using GxE effects with ISLE; in upregulated differentially expressed genes from artery coronary (adjusted- $p = 5.97 \times 10^{-3}$ , adjusted- $p = 9.57 \times 10^{-3}$ ) and artery tibial (adjusted- $p = 1.05 \times 10^{-2}$ , adjusted- $p = 1.55 \times 10^{-2}$ ) tissues using joint effects with both TSLE and DSLE; and in downregulated differentially expressed genes from lung tissue (adjusted- $p = 3.98 \times 10^{-2}$ ) and in up- and down-regulated differentially expressed genes from the spleen (adjusted- $p = 4.71 \times 10^{-2}$ ) using joint effects with ISLE. There was no enrichment using GxE effect with TSLE.

### Post-GWEIS analyses: gene-sets enrichment

Significant gene-sets and GWAS catalogue hits from GWEIS are detailed in Supplementary Tables 11–14, including for UKB *Biocarta*: GPCR pathway; *Reactome*: opioid signalling, neurotransmitter receptor binding and downstream transmission in the postsynaptic cell, transmission across chemical synapses, gastrin CREB signalling pathway via PKC and MAPK; *GWAS catalogue*: post bronchodilator FEV1/FVC ratio, migraine and body mass index. In GS, enrichment was seen using TSLE and DSLE for *GWAS catalogue*: age-related macular degeneration, myopia, urate levels and Heschl's gyrus morphology; and using ISLE for *biological process*: regulation of transporter activity. All adjusted- $p < 0.01$ .

### Cross-cohort prediction

In GS, PRS<sub>D</sub> weighted by the UKB-GWAS of PHQ significantly explained 0.56% of GHQ variance (Empirical- $p < 1.10^{-4}$ ), similar to PRS<sub>MDD</sub> weighted by PGC MDD-GWAS ( $R^2 = 0.78\%$ , Empirical- $p < 1.10^{-4}$ ). PRS<sub>GxE</sub> weighted by the UKB-GWEIS GxE effects explained 0.15% of GHQ variance (Empirical- $p = 0.03$ , Supplementary Table 15). PRS<sub>GxE</sub> fitted jointly with PRS<sub>MDD</sub> significantly improved prediction of GHQ ( $R^2 = 0.93\%$ ,



model  $p = 6.12 \times 10^{-11}$ ; predictive improvement of 19%,  $LRT-p = 5.91 \times 10^{-3}$ ) compared with PRS<sub>MDD</sub> alone. Similar to PRS<sub>GxE</sub> with PRS<sub>D</sub> ( $R^2 = 0.69\%$ , model  $p = 2.72 \times 10^{-8}$ ; predictive improvement of 23%,  $LRT-p = 0.01$ ). PRS<sub>Joint</sub> weighted by the UKB-GWEIS also predicted GHQ ( $R^2 = 0.58\%$ , Empirical- $p < 1.10^{-4}$ ), although



the variance explained was significantly reduced compared with the model fitting  $PRS_{GxE}$  and  $PRS_D$  together ( $LRT-p = 4.69 \times 10^{-7}$ ), suggesting that additive and GxE effects should be modelled independently for polygenic approaches (Fig. 2a).

In UKB (Fig. 2b), both  $PRS_D$  weighted by the GS-GWAS of GHQ and  $PRS_{MDD}$  significantly explained 0.04 and 0.45% of PHQ variance, respectively (both Empirical- $p < 1.10^{-4}$ ; Supplementary Table 15).  $PRS_{GxE}$  derived from the GS-GWEIS GxE effect did not significantly predicted PHQ (TSLE- $PRS_{GxE}$  Empirical- $p = 0.382$ ; DSLE- $PRS_{GxE}$  Empirical- $p = 0.642$ ; ISLE- $PRS_{GxE}$  Empirical- $p = 0.748$ ). Predictive improvements using the  $PRS_{GxE}$  effect fitted jointly with  $PRS_{MDD}$  or  $PRS_D$  were not significant (all  $LRT-p > 0.08$ ).  $PRS_{Joint}$  significantly predicted PHQ (TSLE- $PRS_{Joint}$ :  $R^2 = 0.032\%$ , Empirical- $p < 1.10^{-4}$ ; DSLE- $PRS_{Joint}$ :  $R^2 = 0.012\%$ , Empirical- $p = 4.3 \times 10^{-3}$ ; ISLE- $PRS_{Joint}$ :  $R^2 = 0.032\%$ , Empirical- $p < 1.10^{-4}$ ), although the variance explained was significantly reduced compared with the models fitting  $PRS_{GxE}$  and  $PRS_D$  together (all  $LRT-p < 1.48 \times 10^{-3}$ ).

### Prediction of stress-related traits

Prediction of stress-related traits in independent samples using  $PRS_D$ ,  $PRS_{GxE}$  and  $PRS_{Joint}$  are summarised in Fig. 3a and Supplementary Table 16. Significant trait prediction after FDR adjustment (Empirical- $p < 6.16 \times 10^{-3}$ , FDR-adjusted Empirical- $p < 0.05$ ) using both UKB and GS  $PRS_D$  was seen for: depression status, neuroticism and schizotypal personality.  $PRS_{GxE}$  weighted by the GS-GWEIS GxE effect using ISLE significantly predicted depression status mapped by proxy (Empirical- $p = 7.00 \times 10^{-4}$ , FDR-adjusted Empirical- $p = 9.54 \times 10^{-3}$ ).

Nominally significant predictive improvements ( $LRT-p < 0.05$ ) of fitting  $PRS_{GxE}$  over the  $PRS_D$  effect alone, using summary statistics generated from both UKB and GS were detected for schizotypal personality, heart diseases and chronic obstructive pulmonary disease (COPD) by proxy (Fig. 3b).  $PRS_{GxE}$  weighted by GS-GWEIS GxE effect using ISLE significantly improved prediction over  $PRS_D$  effect alone of depression status mapped by proxy after FDR adjustment ( $LRT-p = 1.96 \times 10^{-4}$ , FDR-adjusted  $LRT-p = 2.35 \times 10^{-2}$ ).

### Discussion

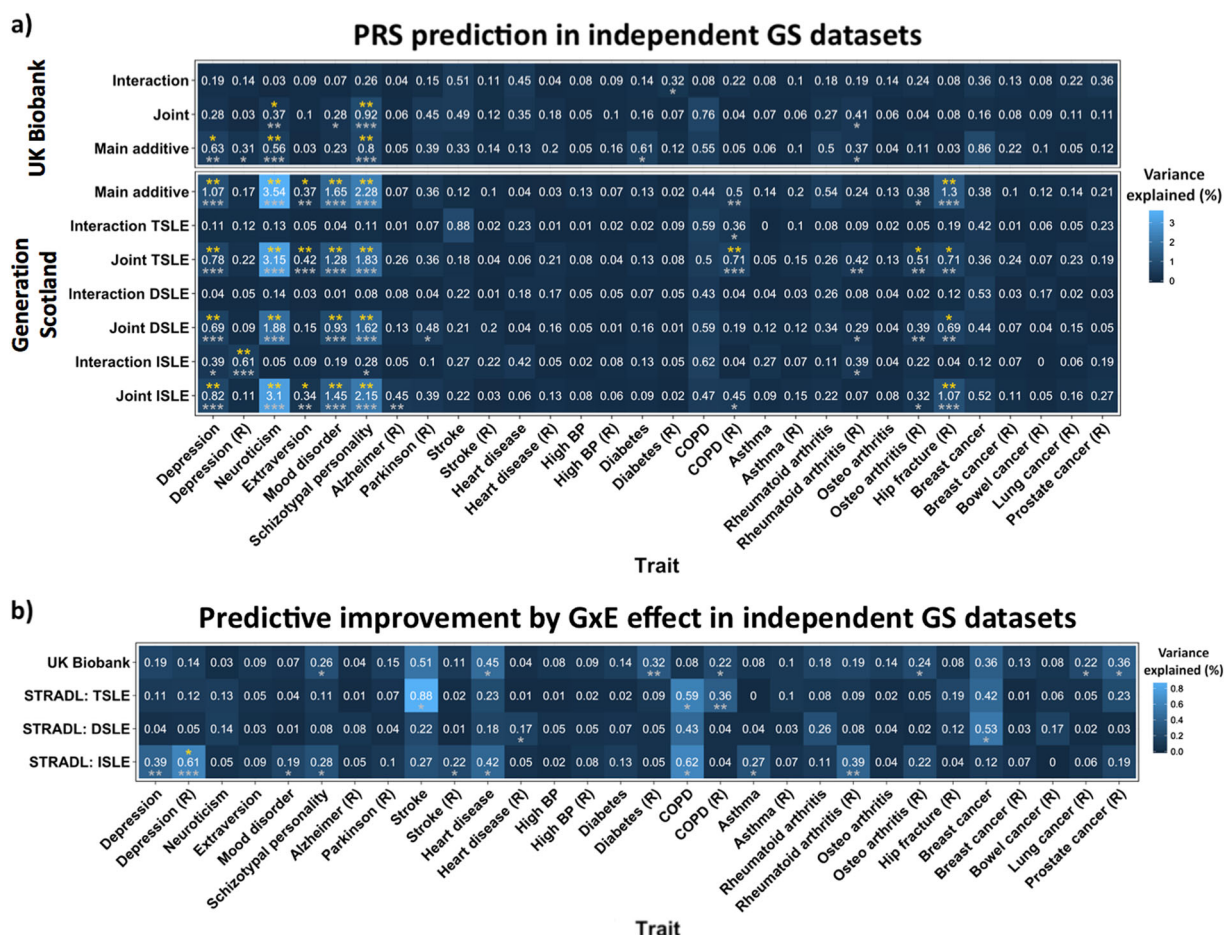
This study performs GWAS and incorporates data on recent adult SLEs into GWEIS of depressive symptoms, identifies new loci and candidate genes for the modulation of genetic response to SLE; and provides insights to help disentangle the underlying aetiological mechanisms increasing the genetic liability through SLE to both depressive symptoms and stress-related traits.

SNP-heritability of depressive symptoms ( $h^2_{SNP} = 9\text{--}13\%$ ), were slightly higher than previous estimates from

African-American populations<sup>34</sup>, and over a third larger than estimates in MDD from European samples<sup>78</sup>.  $h^2_{SNP}$  for PHQ in UKB (9.0%) remained significant after adjusting for SLE (7.9%). Thus, although some genetic contributions may be partially shared between depressive symptoms and reporting of SLE, there is still a relatively large genetic contribution unique to depressive symptoms. Significant  $h^2_{SNP}$  of DSLE in GS (13%) and TSLE<sub>UKB</sub> in UKB (4%), which is mainly composed of dependent SLE items, were detected similar to previous studies (8 and 29%)<sup>34,42</sup>. Conversely, there was no evidence for heritability of independent SLE. A significant bivariate genetic correlation between depressive symptoms and SLE ( $r_G = 0.72$ ) was detected in UKB after adjusting for covariates, suggesting that there are shared common variants underlying self-reported depressive symptoms and SLE. This bivariate genetic correlation was smaller than that estimated from African-American populations ( $r_G = 0.97$ ;  $p = 0.04$ ;  $N = 7179$ )<sup>34</sup>. Genetic correlations between SLE measures and GHQ were not significant in GS ( $N = 4919$ ;  $r_G = 1$ ; all  $p > 0.056$ ), perhaps due to a lack of power in this smaller sample.

Post-GWAS gene-based tests detected six genes significantly associated with PHQ (*DCC*, *ACSS3*, *DRD2*, *STAG1*, *FOXP2* and *KYNU*). Previous studies have implicated these genes in liability to depression (see Supplementary Table 17), and three of them are genome-wide significant in gene-based tests from the latest meta-analysis of major depression that includes UKB (*DCC*,  $p = 2.57 \times 10^{-14}$ ; *DRD2*,  $p = 5.35 \times 10^{-14}$ ; and *KYNU*,  $p = 2.38 \times 10^{-6}$ ;  $N = 807,553$ )<sup>79</sup>. This supports the implementation of quantitative measures such as PHQ to detect genes underlying lifetime depression status<sup>80</sup>. For example, significant gene ontology analysis of the UKB-GWAS identified enrichment for positive regulation of long-term synaptic potentiation, and for previous GWAS findings of brain structure<sup>81</sup>, schizophrenia<sup>82</sup> and response to amphetamines<sup>83</sup>.

The key element of this study was to conduct GWEIS of depressive symptoms and recent SLE. We identified two loci with significant GxE effect in GS. However, none of these associations replicated in UKB ( $p > 0.05$ ). The strongest association was using TSLE at 53-kb downstream of *PIWIL4* (rs12789145). *PIWIL4* is brain expressed and involved in chromatin modification<sup>84</sup>, suggesting it may moderate the effects of stress on depression. It encodes HIWI2, a protein thought to regulate OTX2, that is critical for the development of forebrain and for coordinating critical periods of plasticity disrupting the integration of cortical circuits<sup>85,86</sup>. Indeed, an intronic SNP in *PIWIL4* was identified as the strongest GxE signal in attention deficit hyperactivity disorder using mother's warmth as environmental exposure<sup>87</sup>. The other significant GxE identified in our study was in *ZCCHC2* using



**Fig. 3 Polygenic risk score (PRS) prediction in independent Generation Scotland (GS) datasets.** **a** Heatmap illustrating PRS prediction of a wide range of traits from GS listed in the x axis (Table 1). (R) refers to traits using mapping by proxy approach (i.e., where first-degree relatives of individuals with the disease are considered proxy cases and included into the group of cases). Y axis shows the discovery sample and the effect used to weight PRS. Numbers in cells indicate the % of variance explained, also represented by colour scale. Significance is represented by asterixes according to the following significance codes: \*\* $p < 0.01$ ; \* $p < 0.05$ ; in grey Empirical- $p$ -values after permutation (10,000 times) and in yellow FDR-adjusted Empirical- $p$ -values. **b** Predictive improvement by GxE effect in independent GS datasets. Heatmap illustrating the predictive improvement as a result of incorporating PRS<sub>GxE</sub> into a multiple model along with PRS<sub>D</sub> and covariates (full model), over a model fitting PRS<sub>D</sub> alone with covariates (null model); predicting a wide range of traits from GS listed in the x axis (Table 1). Covariates: age, sex and 20 PCs. (R) refers to traits using mapping by proxy approach (i.e., where first-degree relatives of individuals with the disease are consider proxy cases and included into the group of cases). PRS<sub>GxE</sub> are weighted by genome-wide by environment interaction studies (GWIS) using GxE effects. PRS<sub>D</sub> were weighted by the genome-wide association studies (GWAS) of depressive symptoms additive main effects. The y axis shows the discovery sample used to weight PRS. Numbers in cells indicate the % of variance explained by the PRS<sub>GxE</sub>, also represented by colour scale. Notice that those correspond to the PRS<sub>GxE</sub> predictions in Fig. 3a when PRS<sub>GxE</sub> are weighted by GxE effects. Significance was tested by likelihood ratio tests (LRT): full model including PRS<sub>D</sub> + PRS<sub>GxE</sub> vs. null model with PRS<sub>D</sub> alone (covariates adjusted). Significance is represented by asterixes according to the following significance codes: \*\*\* $p < 0.001$ ; \*\* $p < 0.01$ ; \* $p < 0.05$ ; in grey LRT- $p$ -values and in yellow FDR-adjusted LRT- $p$ -values

DSLE. This zinc-finger protein is expressed in blood CD4 + T cells and is downregulated in individuals with MDD<sup>88</sup> and in those resistant to treatment with citalopram<sup>89</sup>. No GxE effect was seen using ISLE as exposure.

No significant locus or gene with GxE effect was detected in the UKB-GWEIS. Nevertheless, joint effects (the combined additive main and GxE genetic effects) identified two genes significantly associated with PHQ (*ACSS3* and *PHF2*; see Supplementary Table 17). *PHF2*

was recently detected as genome-wide significant at the latest meta-analysis of depression<sup>79</sup>. Notably, *PHF2* paralogs have previously been linked with MDD through stress-response in three other studies<sup>90–92</sup>. Joint effects analyses in GS also detected an additional significant association upstream *CYLC2*, a gene nominally associated ( $p < 1 \times 10^{-5}$ ) with obsessive-compulsive disorder and Tourette's syndrome<sup>93</sup>. Gene-based test from the GS-GWEIS identified a significant association with

*MTNR1B*, a melatonin receptor gene, using DSLE (both GxE and joint effect; Supplementary Table 17). Genes prioritised using GxE effects were enriched in differentially expressed genes from several tissues including the adrenal gland, which releases cortisol into the bloodstream in response to stress, thus playing a key role in the stress-response system, reinforcing a potential role of GxE in stress-related conditions.

The different instruments and sample sizes available make it hard to compare results between cohorts. Whereas GS contains deeper phenotyping measurements of stress and depressive symptoms than UKB, the sample size is much smaller, which may be reflected in the statistical power required to reliably detect GxE effects. Furthermore, the presence and size of GxE effects are dependent on their parameterisation (i.e., the measurement, scale and distribution of the instruments used to test such interaction)<sup>94</sup>. Thus, GxE may be incomparable across GWEIS due to differences in both phenotype assessment and stressors tested. Although our results suggest that both depressive symptom measures are correlated with lifetime depression status, different influences on depressive symptoms from the SLE covered across studies may contribute to lack of stronger replication. Instruments in GS cover a wider range of SLE and are more likely to capture changes in depressive symptoms as consequence of their short-term effects. Conversely, UKB could capture more marked long-term effects, as SLE were captured over 2 years compared with the 6 months in GS. New mental health questionnaires covering a wide range of psychiatric symptoms and SLE in the latest release of UKB data provides the opportunity to create similar measures to GS in the near future. Further replication in independent studies with equivalent instruments is required to validate our GWEIS findings. Despite these limitations and a lack of overlap in the individual genes prioritised from the two GWEIS, replication was seen in the predictive improvement of using PRS<sub>GxE</sub> derived from the GWEIS GxE effects to predict stress-related phenotypes.

The third aim of this study was to test whether modelling GxE effects could improve predictive genetic models, and thus help to explain deviation from additive models and missing heritability for MDD<sup>95</sup>. Multiple regression models suggested that inclusion of PRS<sub>GxE</sub> weighted by GxE effects could improve prediction of an individual's depressive symptoms over use of PRS<sub>MDD</sub> or PRS<sub>D</sub> weighted by additive effects alone. In GS, we detected a predictive gain of 19% over PRS<sub>MDD</sub> weighted by PGC MDD-GWAS, and a gain of 23% over PRS<sub>D</sub> weighted by UKB-GWAS (Fig. 2a). However, these findings did not surpass stringent Bonferroni correction and could not be validated in UKB. This may reflect in the poor predictive power of the PRS generated from the

much smaller GS discovery sample. The results show a noticeably reduced prediction using PRS<sub>Joint</sub> weighted by joint effects, which suggests that the genetic architecture of stress-response is at least partially independent and differs from genetic additive main effects. Overall, our results from multiple regression models suggest that for polygenic approaches main and GxE effects should be modelled independently.

SLE effects are not limited to mental illness<sup>45</sup>. Our final aim was to investigate shared aetiology between GxE for depressive symptoms and stress-related traits. Despite the differences between the respondents and non-respondents (Table 1 legend), a significant improvement was seen in predicting depressive status when mapping by proxy cases using GxE effect from GS-GWEIS with independent SLE (*FDR-adjusted LRT-p* = 0.013), but not with dependent SLE. GxE effects using statistics generated from both discovery samples, despite the differences in measures, nominally improved the phenotypic prediction of schizotypal personality, heart disease and the proxy of COPD (*LRT-p* < 0.05). Other studies have also found evidence supporting a link between stress and depression in these phenotypes (see Supplementary Material for extended review) and suggest, for instance, potential pleiotropy between schizotypal personality and stress-response. Our findings point to a potential genetic component underlying a stress-response-depression-comorbidities link due, at least in part, to shared stress-response mechanisms. A relationship between SLE, depression and coping strategies such as smoking suggests that genetic stress-response may modulate adaptive behaviours such as smoking, fatty diet intake, alcohol consumption and substance abuse. This is discussed further in the Supplementary Material.

In this study, evidence for SNPs with significant GxE effects came primarily from the analyses of dependent SLE and not from independent SLE. This supports a genetic effect on probability of exposure to, or reporting of SLE, endorsing a gene–environment correlation. Chronic stress may influence cognition, decision making and behaviour eventually leading to higher risk taking<sup>96</sup>. These conditions may also increase sensitivity to stress among vulnerable individuals, including those with depression, who also have a higher propensity to report SLE, particularly dependent SLE<sup>38</sup>. A potential reporting bias in dependent SLE may be mediated as well by heritable behavioural, anxiety or psychological traits such as risk taking<sup>42,97</sup>. Furthermore, individuals vulnerable to MDD may behave in a manner that exposes them more frequently to high risk and stressful environments<sup>14</sup>. This complex interplay, reflected in the form of a gene–environment correlation effect, would hinder the interpretation of GxE effects from GWEIS as pure interactions. A mediation of associations between SLE and



depressive symptoms, through genetically driven sensitivity to stress, personality or behavioural traits would support the possibility of subtle genotype-by-genotype (GxG) interactions, or genotype-by-genotype-by-environment (GxGxE) interactions, contributing to depression<sup>98,99</sup>. In contrast, PRS prediction of the stress-related traits: schizotypal personality, heart disease and COPD, was primarily from derived weights using independent SLE, suggesting that a common set of variants moderate the effects of SLE across stress-related traits and that larger sample sizes will be required to detect the individual SNPs contributing to this. Thus, our findings support the inclusion of environmental information into GWAS to enhance the detection of relevant genes. The results of studying dependent and independent SLE support a contribution of genetically mediated exposure to and/or reporting of SLE, perhaps through sensitivity to stress as mediator.

This study emphasises the relevance of GxE in depression and human health in general and provides the basis for future lines of research.

# Acknowledgements

See Supplementary Material: acknowledgements.

# Author details

<sup>1</sup>Medical Genetics Section, University of Edinburgh, Centre for Genomic and Experimental Medicine and MRC Institute of Genetics and Molecular Medicine, Edinburgh, UK. <sup>2</sup>Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, UK. <sup>3</sup>Division of Psychiatry, Deanery of Clinical Sciences, University of Edinburgh, Royal Edinburgh Hospital, Morningside Park, Edinburgh EH10 5HF, UK. <sup>4</sup>Health Informatics Centre, University of Dundee, Dundee, UK. <sup>5</sup>Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

# Conflict of interest

The authors declare that they have no conflict of interest.

# Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Supplementary information** accompanies this paper at (<https://doi.org/10.1038/s41398-018-0360-y>).

Received: 26 November 2018 Accepted: 10 December 2018

Published online: 17 January 2019

# References

- Plomin, R., Owen, M. & McGuffin, P. The genetic basis of complex human behaviors. *Science* **264**, 1733–1739 (1994).
- Kendler, K. S. & Eaves, L. J. Models for the joint effect of genotype and environment on liability to psychiatric illness. *Am. J. Psychiatry* **143**, 279–289 (1986).
- Kendler, K. S. & Baker, J. H. Genetic influences on measures of the environment: a systematic review. *Psychol. Med.* **37**, 615–626 (2007).
- Colodro-Conde, L. et al. A direct test of the diathesis-stress model for depression. *Mol. Psychiatry* **23**, 1590–1596 (2017).
- Luciano, M. et al. Shared genetic aetiology between cognitive ability and cardiovascular disease risk factors: Generation Scotland's Scottish family health study. *Intelligence* **38**, 304–313 (2010).
- Kessler, R. C. & Bromet, E. J. The epidemiology of depression across cultures. *Annu. Rev. Public Health* **34**, 119–138 (2013).
- Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: review and meta-analysis. *Am. J. Psychiatry* **157**, 1552–1562 (2000).
- Hammen, C. Stress and depression. *Annu. Rev. Clin. Psychol.* **1**, 293–319 (2005).
- Kessler, R. C. The effects of stressful life events on depression. *Annu. Rev. Psychol.* **48**, 191–214 (1997).
- Kendler, K. S., Karkowski, L. M. & Prescott, C. A. Causal relationship between stressful life events and the onset of major depression. *Am. J. Psychiatry* **156**, 837–841 (1999).
- Stroud, C. B., Davila, J. & Moyer, A. The relationship between stress and depression in first onsets versus recurrences: a meta-analytic review. *J. Abnorm. Psychol.* **117**, 206–213 (2008).
- Kendler, K. S., Karkowski, L. M. & Prescott, C. A. Stressful life events and major depression: risk period, long-term contextual threat, and diagnostic specificity. *J. Nerv. Ment. Dis.* **186**, 661–669 (1998).
- Silberg, J., Rutter, M., Neale, M. & Eaves, L. Genetic moderation of environmental risk for depression and anxiety in adolescent girls. *Br. J. Psychiatry* **179**, 116–121 (2001).
- Kendler, K. S. et al. Stressful life events, genetic liability, and onset of an episode of major depression in women. *Am. J. Psychiatry* **152**, 833–842 (1995).
- Arnaud-Soler, A., Adams, M., Hayward, C. & Thomson, P. Genome-wide interaction study of a proxy for stress-sensitivity and its prediction of major depressive disorder. *PLoS One* **13**, e0209160 (2018).
- Arnaud-Soler, A. et al. A validation of the diathesis-stress model for depression in Generation Scotland. *Transl. Psychiatry* (in press).
- Garantziotis, S. & Schwartz, D. A. Ecogenomics of respiratory diseases of public health significance. *Annu. Rev. Public Health* **31**, 37–51 (2010).
- Aschard, H. et al. Evidence for large-scale gene-by-smoking interaction effects on pulmonary function. *Int. J. Epidemiol.* **46**, 894–904 (2017).
- Molfini, N. A. & Coyle, A. J. Gene-environment interactions in chronic obstructive pulmonary disease. *Int. J. Chron. Obstruct. Pulmon. Dis.* **3**, 491–497 (2008).
- Polonikov, A. V., Ivanov, V. P. & Solodilova, M. A. Genetic variation of genes for xenobiotic-metabolizing enzymes and risk of bronchial asthma: the importance of gene-gene and gene-environment interactions for disease susceptibility. *J. Hum. Genet.* **54**, 440–449 (2009).
- Haiman, C. A. et al. Ethnic and racial differences in the smoking-related risk of lung cancer. *N. Engl. J. Med.* **354**, 333–342 (2006).
- Han, J., Hankinson, S. E., Colditz, G. A. & Hunter, D. J. Genetic variation in XRCC1, sun exposure, and risk of skin cancer. *Br. J. Cancer* **91**, 1604–1609 (2004).
- Manning, A. K. et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44**, 659–669 (2012).
- Wang, L., Murk, W. & DeWan, A. T. Genome-wide gene by environment interaction analysis identifies common SNPs at 17q21.2 that are associated with increased body mass index only among asthmatics. *PLoS One* **10**, e0144114 (2015).
- Siebert, S. et al. Genome-wide investigation of gene-environment interactions in colorectal cancer. *Hum. Genet.* **132**, 219–231 (2013).
- Gong, J. et al. Genome-wide interaction analyses between genetic variants and alcohol consumption and smoking for risk of colorectal cancer. *PLoS. Genet.* **12**, e1006296 (2016).
- Polfus, L. M. et al. Genome-wide association study of gene by smoking interactions in coronary artery calcification. *PLoS One* **8**, e74642 (2013).
- Duncan, L. E. & Keller, M. C. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *Am. J. Psychiatry* **168**, 1041–1049 (2011).
- Risch, N. et al. Interaction between the serotonin transporter gene (5-HTTLPR), stressful life events, and risk of depression: a meta-analysis. *JAMA* **301**, 2462–2471 (2009).
- Karg, K., Burmeister, M., Shedden, K. & Sen, S. The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: evidence of genetic moderation. *Arch. Gen. Psychiatry* **68**, 444–454 (2011).
- Bleys, D., Luyten, P., Soenens, B. & Claes, S. Gene-environment interactions between stress and 5-HTTLPR in depression: a meta-analytic update. *J. Affect Disord.* **226**, 339–345 (2018).

32. Peyrot, W. J. et al. Does childhood trauma moderate polygenic risk for depression? A meta-analysis of 5765 subjects from the psychiatric genomics consortium. *Biol. Psychiatry* **84**, 138–147 (2018).
33. Kraft, P., Yen, Y. C., Stram, D. O., Morrison, J. & Gauderman, W. J. Exploiting gene-environment interaction to detect genetic associations. *Hum. Hered.* **63**, 111–119 (2007).
34. Dunn, E. C. et al. Genome-wide association study (GWAS) and genome-wide by environment interaction study (GWEIS) of depressive symptoms in African American and Hispanic/Latina women. *Depress. Anxiety* **33**, 265–280 (2016).
35. Otowa, T. et al. The first pilot genome-wide gene-environment study of depression in the Japanese population. *PLoS One* **11**, e0160823 (2016).
36. Ikeda, M. et al. Genome-wide environment interaction between depressive state and stressful life events. *J. Clin. Psychiatry* **77**, e29–e30 (2016).
37. Plomin, R., DeFries, J. C. & Loehlin, J. C. Genotype-environment interaction and correlation in the analysis of human behavior. *Psychol. Bull.* **84**, 309–322 (1977).
38. Clarke, T. et al. Genetic and environmental determinants of stressful life events and their overlap with depression and neuroticism [version 1; referees: 3 approved with reservations]. *Wellcome Open Res.* **3**, 11 (2019).
39. Brugha, T., Bebbington, P., Tennant, C. & Hurry, J. The list of threatening experiences: a subset of 12 life event categories with considerable long-term contextual threat. *Psychol. Med.* **15**, 189–194 (1985).
40. Kendler, K. S., Karkowski, L. M. & Prescott, C. A. The assessment of dependence in the study of stressful life events: validation using a twin design. *Psychol. Med.* **29**, 1455–1460 (1999).
41. Plomin, R., Lichtenstein, P., Pedersen, N. L., McClearn, G. E. & Nesselroade, J. R. Genetic influence on life events during the last half of the life span. *Psychol. Aging* **5**, 25–30 (1990).
42. Power, R. A. et al. Estimating the heritability of reporting stressful life events captured by common genetic variants. *Psychol. Med.* **43**, 1965–1971 (2013).
43. Bemmels, H. R., Burt, S. A., Legrand, L. N., Iacono, W. G. & McGue, M. The heritability of life events: an adolescent twin and adoption study. *Twin. Res. Hum. Genet.* **11**, 257–265 (2008).
44. Boardman, J. D., Alexander, K. B. & Stallings, M. C. Stressful life events and depression among adolescent twin pairs. *Biodemography Soc. Biol.* **57**, 53–66 (2011).
45. Salleh, M. R. Life event, stress and illness. *Malays. J. Med. Sci.* **15**, 9–18 (2008).
46. Thaipisutikul, P., Ittasakul, P., Waleprakhon, P., Wisajun, P. & Julagat, S. Psychiatric comorbidities in patients with major depressive disorder. *Neuropsychiatr. Dis. Treat.* **10**, 2097–2103 (2014).
47. Moussavi, S. et al. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet* **370**, 851–858 (2007).
48. Topic, R. et al. Somatic comorbidity, metabolic syndrome, cardiovascular risk, and CRP in patients with recurrent depressive disorders. *Croat. Med. J.* **54**, 453–459 (2013).
49. Lloyd, C. E., Roy, T., Nouwen, A. & Chauhan, A. M. Epidemiology of depression in diabetes: international and cross-cultural issues. *J. Affect. Disord.* **142**(Suppl), S22–S29 (2012).
50. Ohayon, M. M. & Schatzberg, A. F. Using chronic pain to predict depressive morbidity in the general population. *Arch. Gen. Psychiatry* **60**, 39–47 (2003).
51. Slavich, G. M. & Irwin, M. R. From stress to inflammation and major depressive disorder: a social signal transduction theory of depression. *Psychol. Bull.* **140**, 774–815 (2014).
52. Smith, B. H. et al. Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int. J. Epidemiol.* **42**, 689–700 (2013).
53. Gunderson, K. L. Whole-genome genotyping on bead arrays. in *DNA Microarrays for Biomedical Research: Methods and Protocols* (ed. Dufva, M.) 197–213 (Humana Press, Totowa, NJ, 2009).
54. Kerr, S. M. et al. Pedigree and genotyping quality analyses of over 10,000 DNA samples from the Generation Scotland: Scottish Family Health Study. *Bmc. Med. Genet.* **14**, 38 (2013).
55. Nagy, R. et al. Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants. *Genome Med.* **9**, 23 (2017).
56. Navrady, L. B. et al. Cohort profile: stratifying resilience and depression longitudinally (STRADL): a questionnaire follow-up of Generation Scotland: Scottish Family Health Study (GS:SFHS). *Int. J. Epidemiol.* **47**, 13–14g (2018).
57. Smith, B. H. et al. Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *Bmc. Med. Genet.* **7**, 74 (2006).
58. Fernandez-Pujals, A. M. et al. Epidemiology and heritability of major depressive disorder, stratified by age of onset, sex, and illness course in Generation Scotland: Scottish Family Health Study (GS:SFHS). *PLoS One* **10**, e0142197 (2015).
59. Huang, J. et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat. Commun.* **6**, 8111 (2015).
60. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS. Med.* **12**, e1001779 (2015).
61. Protocol for a large-scale prospective epidemiological resource. *UK Biobank* (2006) [www.ukbiobank.ac.uk/resources/](http://www.ukbiobank.ac.uk/resources/). [www.ukbiobank.ac.uk/wp-content/uploads/2011/11/UK-Biobank-Protocol.pdf](http://www.ukbiobank.ac.uk/wp-content/uploads/2011/11/UK-Biobank-Protocol.pdf).
62. UK Biobank Ethics and Governance Framework, Version 3.0. *UK Biobank* (2007). Ethics and Governance Framework. [www.ukbiobank.ac.uk/resources/](http://www.ukbiobank.ac.uk/resources/). [www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf](http://www.ukbiobank.ac.uk/wp-content/uploads/2011/05/EGF20082.pdf).
63. Brugha, T. S. & Cragg, D. The list of threatening experiences: the reliability and validity of a brief life events questionnaire. *Acta Psychiatr. Scand.* **82**, 77–81 (1990).
64. Motrico, E. et al. Psychometric properties of the list of threatening experiences—LTE and its association with psychosocial factors and mental disorders according to different scoring methods. *J. Affect. Disord.* **150**, 931–940 (2013).
65. Goldberg, D. P. & Hillier, V. F. A scaled version of the general health questionnaire. *Psychol. Med.* **9**, 139–145 (1979).
66. Sterling, M. General health questionnaire — 28 (GHQ-28). *J. Physiother.* **57**, 259 (2011).
67. Goldberg, D. P. et al. The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychol. Med.* **27**, 191–197 (1997).
68. Wang, L. et al. [Value of patient health questionnaires (PHQ)—9 and PHQ-2 for screening depression disorders in cardiovascular outpatients]. *Zhonghua Xin Xue Guan Bing. Za Zhi* **43**, 428–431 (2015).
69. Spitzer, R. L., Kroenke, K. & Williams, J. B. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. Patient health questionnaire. *JAMA* **282**, 1737–1744 (1999).
70. Smith, D. J. et al. Prevalence and characteristics of probable major depression and bipolar disorder within UK biobank: cross-sectional study of 172,751 participants. *PLoS One* **8**, e75362 (2013).
71. Liu, J. Z., Erlich, Y. & Pickrell, J. K. Case-control association mapping by proxy using family history of disease. *Nat. Genet.* **49**, 325–331 (2017).
72. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
73. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
74. Almlil, L. M. et al. Correcting systematic inflation in genetic association tests that consider interaction effects: application to a genome-wide association study of posttraumatic stress disorder. *JAMA Psychiatry* **71**, 1392–1399 (2014).
75. Keller, M. C. Gene x environment interaction studies have not properly controlled for potential confounders: the problem and the (simple) solution. *Biol. Psychiatry* **75**, 18–24 (2014).
76. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. FUMA: functional mapping and annotation of genetic associations. *Nat. Commun.* **8**, 1826 (2017).
77. Euesden, J., Lewis, C. M. & O'Reilly, P. F. PRSice: polygenic risk score software. *Bioinformatics* **31**, 1466–1468 (2015).
78. Cross-Disorder Group of the Psychiatric Genomics et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
79. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Commun.* (in press).
80. Altman, D. G. & Royston, P. The cost of dichotomising continuous variables. *BMJ* **332**, 1080 (2006).
81. Stein, J. L. et al. Voxelwise genome-wide association study (vGWAS). *Neuroimage* **53**, 1160–1174 (2010).
82. Li, Z. et al. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nat. Genet.* **49**, 1576–1583 (2017).
83. Hart, A. B. et al. Genome-wide association study of d-amphetamine response in healthy volunteers identifies putative associations, including cadherin 13 (CDH13). *PLoS One* **7**, e42646 (2012).
84. Sugimoto, K. et al. The induction of H3K9 methylation by PIWIL4 at the p16lnk4a locus. *Biochem. Biophys. Res. Commun.* **359**, 497–502 (2007).
85. Sivagurunathan, S., Arunachalam, J. P. & Chidambaram, S. PIW-like protein, HIW2 is aberrantly expressed in retinoblastoma cells and affects cell-cycle potentially through OTX2. *Cell. Mol. Biol. Lett.* **22**, 17 (2017).



86. Lee, H. H. C. et al. Genetic Otx2 mis-localization delays critical period plasticity across brain regions. *Mol. Psychiatry* **22**, 680–688 (2017).
87. Sonuga-Barke, E. J. et al. Does parental expressed emotion moderate genetic effects in ADHD? An exploration using a genome wide association scan. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **147B**, 1359–1368 (2008).
88. Belzeaux, R. et al. Responder and nonresponder patients exhibit different peripheral transcriptional signatures during major depressive episode. *Transl. Psychiatry* **2**, e185 (2012).
89. Mamdani, F., Berlim, M. T., Beaulieu, M. M. & Turecki, G. Pharmacogenomic predictors of citalopram treatment outcome in major depressive disorder. *World J. Biol. Psychiatry* **15**, 135–144 (2014).
90. Wong, M. L. et al. The PHF21B gene is associated with major depression and modulates the stress response. *Mol. Psychiatry* **22**, 1015–1025 (2017).
91. Walsh, R. M. et al. Phf8 loss confers resistance to depression-like and anxiety-like behaviors in mice. *Nat. Commun.* **8**, 15142 (2017).
92. Shi, G. et al. PHD finger protein 2 (PHF2) represses ribosomal RNA gene transcription by antagonizing PHF finger protein 8 (PHF8) and recruiting methyltransferase SUV39H1. *J. Biol. Chem.* **289**, 29691–29700 (2014).
93. Yu, D. et al. Cross-disorder genome-wide analyses suggest a complex genetic relationship between Tourette's syndrome and OCD. *Am. J. Psychiatry* **172**, 82–93 (2015).
94. Eaves, L. J., Last, K., Martin, N. G. & Jinks, J. L. A progressive approach to non-additivity and genotype-environmental covariance in the analysis of human differences. *Br. J. Math. Stat. Psychol.* **30**, 1–42 (1977).
95. Polderman, T. J. et al. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat. Genet.* **47**, 702–709 (2015).
96. Ceccato, S., Kudielka, B. M. & Schwieren, C. Increased risk taking in relation to chronic stress in adults. *Front. Psychol.* **6**, 2036 (2015).
97. Kandler, C., Bleidorn, W., Riemann, R., Angleitner, A. & Spinath, F. M. Life events as environmental states and genetic traits and the role of personality: a longitudinal twin study. *Behav. Genet.* **42**, 57–72 (2012).
98. Conway, C. C., Hammen, C., Brennan, P. A., Lind, P. A. & Najman, J. M. Interaction of chronic stress with serotonin transporter and catechol-O-methyltransferase polymorphisms in predicting youth depression. *Depress Anxiety* **27**, 737–745 (2010).
99. Cicchetti, D. & Rogosch, F. A. Genetic moderation of child maltreatment effects on depression and internalizing symptoms by serotonin transporter linked polymorphic region (5-HTTLPR), brain-derived neurotrophic factor (BDNF), norepinephrine transporter (NET), and corticotropin releasing hormone receptor 1 (CRHR1) genes in African American children. *Dev. Psychopathol.* **26**, 1219–1239 (2014).